The Role of Weighting in Econometric Analysis

Annett Seibold – Chair in Macroeconomics – JGU Mainz

October 2025

1 Challenges of Weighting in Econometrics

Solon et al. (2015) argue that weighting is not a technique that is used for a single purpose. It is necessary to think about the specific purpose before applying weights. There are two broad goals: the estimation of descriptive statistics and the estimation of causal effects. When the given sample is representative for the target population, which you can receive for example through a "simple random sample draw" (Solon et al., 2015, p. 304), weighting is not required to ensure representativeness.

For descriptive statistics, the process is relatively straight forward. The aim is to get representative estimates for a target group given a sample from the target group. When this sample is not representative for the target group, weights need to be applied. The weights are the inverse probability of selection to correct imbalances (Solon et al., 2015). Most of the time, the weights are already included in the data set and are available for use. They are scaled to approximate the total size of the target population. If we restrict the dataset to a subsample, the weights are not automatically representative for the subsample. So, the weights need to be recalibrated for the subpopulation to correctly represent the target subgroup if the goal is to correctly extrapolate the subgroup. This can be done via poststratification or ranking to match the weights (Valliant and Dever, 2018). If this is not the goal, a recalibration of the weights is not necessary.

For causal effects, the situation is more complex. In this case, Solon et al. (2015) identity three motives for weighting: correcting for heteroskedasticity, correcting in the presence of endogenous sampling and identifying average partial effects when causal impacts are heterogenous and not modelled.

Concerning the first motive, weighting is often used to achieve more precise and efficient results by addressing heteroskedasticity (non-constant error variance) in regressions. Contrary, Dickens (1990) showed that with weighting standard errors are increased compared to the unweighted OLS if the standard errors are clustered. Solon et al. (2015) therefore suggest that the best practice is to use robust standard errors and to conduct standard heteroskedasticity diagnostics (e.g. the modified Breusch-Pagan test). This test helps to assess whether heteroskedasticity is present and reports heteroskedasticity-robust standard errors and compares OLS and WLS estimates. Differences in the coefficients show misspecifications or endogenous sampling. When the model is correctly specified OLS and WLS are consistent.

Second, weighting can be used to correct for endogenous sampling which means that the probability of being in the sample depends on the outcome variable. Ignoring this will usually lead to bias in the regressions. Weighting by the inverse probability of selection can fix this bias. This occurs for example in choice-based sampling or a survey of a subgroup of the population. If sampling differences do not occur based on the outcome variable, weighting may reduce precision. The best practice according to Solon et al. (2015) is to use inverse probability weights and robust standard errors if sampling is endogenous. If sampling is exogenous, OLS

and WLS are unbiased, but OLS is often more precise. Both results should be reported for comparison.

Third, weighting is often used to identify average partial effects when effects differ across people. If the effect in interest is different for different groups (heterogenous effects) and you don't include that variation in your model, weighting will usually not give the true population average effect. Solon et al. (2015) state as a best practice not to use weighting to fix this problem and that weighted and unweighted results should be compared. Expected heterogeneity should be addressed directly rather than trying to use weighting.

In conclusion, an overall best practice is to clearly state your reason for weighting, verify the applicability of that reason and report both weighted and unweighted estimates. Robust standard errors should always be used to guard against misspecified variance structures.

2 Weighting in Stata

Stata provides a framework for applying weights in statistical estimation. According to the *Stata 19 User's Guide* (StataCorp 2025), four distinctive types of weights are available: frequency weights, analytic weights, probability weights and importance weights. Each type has different implication for analysis and depends on the type of data available.

Frequency weights (*fweight*) are the most straightforward. They indicate replication counts which means that a single observation in the dataset actually represents multiple identical cases. An observation with *fweight* of 5 is equivalent to the same observation appearing 5 times in the dataset. *Fweight* is important for data handling when working with collapsed or aggregated data but is not interesting from a statistical perspective.

Analytical weights (*aweight*) could also be called precision weights and are used when the dataset contains averages rather than individual observations. For example, you need *aweight* for a linear regression on data which are observed means. In this case, the weights are inversely proportional to the variance of each observation. So, larger weights indicate more precise estimates (eg. means based on larger groups). Stata normalizes the calculations to sum to *N* and then uses the weights as if they were *fweight*.

Probability weights or sampling weights (*pweight*) are the appropriate tool for survey data. They resemble the inverse of the probability of an observation being included on the sample. In other words, each case in the dataset represents a certain number of individuals in the population. Stata automatically uses robust variance estimation when using *pweight*. If the survey design involves stratification or clustering, this should be explicitly declared using the *svy* commands. This guarantees that both point estimates and their standard errors reflect the survey design.

Finally, importance weights (*iweight*) are a more general category and are used primarily by programmers to control the influence of particular observations. They are rarely relevant for applied statistical analysis.

For the work with SOEP, probability weights should be used whenever possible. In the dataset, *phrf_1* represents the probability weight, *hid* is the primary sampling unit (PSU) and *psample* is the stratification variable. This way, Stata will correctly account for the survey design and produce estimates of population parameters with robust standard errors.

3 Literature

Dickens, W. T. (1990). Error Components in Grouped Data: Is It Ever Worth Weighting? *Review of Economics and Statistics, 72*(2), 328–33.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For? *The Journal of Human Resources*, *50*(2), 301–316.

StataCorp (2025). *Stata 19 User's Guide*. Stata Press.

Valliant, R., & Dever, J. A. (2018). *Survey Weights: A Step-by-Step Guide to calculation*. Stata Press.